

AMENDMENTS TO THE CLAIMS:

1. (Currently amended) An apparatus for encoding a DNA sequence to achieve a high data compression ratio for storage or transfer, which comprises:

a comparative unit for aligning a reference sequence having known DNA information with a subject sequence to be encoded-compressed and extracting a difference between the reference sequence and the subject sequence;

a conversion unit for converting information of the extracted difference between the reference sequence and the subject sequence into a string of predetermined characters and for outputting the string of characters;

a code storage unit for storing predetermined a conversion codes code that corresponds to the individual characters a character to represent the extracted difference; and

an encoding unit for encoding the individual characters that make the string of the characters using the conversion codes code.

2. (Currently amended) The apparatus of claim 1, wherein the characters to represent the extracted difference comprisescomprise

a first character representing each DNA base-symbols,

a second-numeric character representing the-a number of base positions that characterize a feature of the extracted difference,

a third character representing the-starting and-or ending of the extracted difference, and

a fourth character representing continuation of the-whether a type of extracted difference occurs in succession in the subject sequence.

3. (Currently amended) The apparatus of claim 2, wherein the-conversion-unit-converts respective information of-features of the extracted difference converted into the string of characters comprise

starting of the extracted difference,

a start position of the extracted difference,

continuation-whether a type of extracted difference occurs in succession in the subject sequence, the

a number of continued bases in the extracted difference,

a basesbase which the extracted difference comprises,

ending of the extracted difference, and

a distance between the start position and the end position of the extracted difference into the third character, the second character, the fourth character, the second character, the first character, the third character, and the second character, and outputs the string of the characters.

4. (Currently amended) The apparatus of claim 1, wherein a type of the extracted difference comprises

a start region mismatch between the reference sequence and the subject sequence;
a blank representing there is no base in a base position in the subject sequence by base deletion of the subject sequence corresponding to the reference sequence;
a single base pair mismatch between the reference sequence and the subject sequence;
a base insertion into the subject sequence;
a multiple base pair mismatch between the reference sequence and the subject sequence; and
or
an end region mismatch between the reference sequence and the subject reference.

5. (Currently amended) The apparatus of claim 1, wherein the conversion codes are 4 bit codes, each of which corresponds to each of the characters.

6. (Currently amended) The apparatus of claim 1, which further comprises
a division unit for dividing the extracted difference into segments of a predetermined size, and
wherein the conversion unit converts information of the extracted difference into the string which is made up of the characters to represent the extracted difference based on the segments.

7. (Currently amended) The apparatus of claim 1, which further comprises:
a compression unit for compressing the encoded subject sequence; and
a sequence storage unit for storing the compressed subject sequence.

8. (Currently amended) The apparatus of claim 1, which further comprises
a pre-processing unit for modifying the reference sequence using creating a variation sequence generation factor from created by a variation sequence generation function that uses random variables as inputs and modifying the reference sequence using the created variation sequence generation factor.

9. (Currently amended) The apparatus of claim 8, wherein the variation sequence induction generation factor comprises ~~the~~ a total number of variations, a distance between ~~the two adjacent~~ variations, a length of ~~the~~ a variationsvariation, a type of the variationsvariation, and a variation sequence of the variation.

10. (Withdrawn) A method for encoding a DNA sequence, which comprises:
aligning a reference sequence having known DNA information with a subject sequence to be encoded;
extracting a difference between the reference sequence and the subject sequence;
converting information of the extracted difference between the reference sequence and the subject sequence into a string of predetermined characters; and
encoding the individual characters that make the string of the predetermined characters using predetermined conversion codes that correspond to the individual characters.

11. (Withdrawn) The method of claim 10, wherein the characters comprises a first character representing DNA base symbols, a second character representing the number of the difference, a third character representing the starting and ending of the difference, and a fourth character representing continuation of the difference.

12. (Withdrawn) The method of claim 11, wherein converting comprises:
allotting the third character for the starting of the difference;
allotting the second character for the starting position of the difference;
allotting the fourth character for the continuation of the difference;
allotting the second character for the number of the continued bases of the difference;
allotting the first character for the bases of the difference;
allotting the third character for the ending of the difference;
allotting the second character for the distance between the start position and the end position of the difference; and
outputting the string of the allotted characters.

13. (Withdrawn) The method of claim 10, wherein the difference comprises start region mismatch between the reference sequence and the subject sequence, blank by base deletion of the subject sequence corresponding to the reference sequence, single base pair mismatch between the reference sequence and the subject sequence, base insertion into the subject sequence, multiple base

pair mismatch between the reference sequence and the subject sequence, and end region mismatch between the reference sequence and the subject reference.

14. (Withdrawn) The method of claim 10, wherein the conversion codes are 4 bit codes, each of which corresponds to each of the characters.

15. (Withdrawn) The method of claim 10, which further comprises dividing the extracted difference into segments of predetermined sizes, and
wherein in converting, information of the extracted difference is converted into the string of the characters based on the segments.

16. (Withdrawn) The method of claim 10, which further comprises:
compressing the encoded subject sequence; and
storing the compressed subject sequence.

17. (Withdrawn) The method of claim 10, which further comprises, before aligning, creating a variation sequence induction factor from a variation sequence induction function that uses random variables as inputs and modifying the reference sequence using the created variation sequence induction factor.

18. (Withdrawn) The method of claim 17, wherein the variation sequence induction factor comprises the total number of variations, distance between the variations, length of the variations, type of the variations, and a variation sequence.

19. (Currently amended) A computer readable medium having embodied thereon a computer program for a method for encoding a DNA sequence to achieve a high data compression ratio, the method comprising:

aligning a reference sequence having known DNA information with a subject sequence to be encoded;

extracting a difference between the reference sequence and the subject sequence;

converting ~~information of~~ the extracted difference between the reference sequence and the subject sequence into a string of ~~predetermined~~ characters; and

encoding ~~the individual characters that make the string of the characters using predetermined~~
a conversion codes code that corresponds to the a individual characters,

wherein the computer readable medium is not a carrier wave.